

## **СЪВРЕМЕНО СЪСТОЯНИЕ НА БЪЛГАРСКО-АНГЛИЙСКИЯ ПАРАЛЕЛЕН КОРПУС СЪС СЪОТНЕСЕНИ ГЛАГОЛНИ ФОРМИ**

**ТОДОР ЛАЗАРОВ**

ИНСТИТУТ ЗА БЪЛГАРСКИ ЕЗИК „ПРОФ. ЛЮБОМИР АНДРЕЙЧИН“ ПРИ БАН  
*tdlazarov@gmail.com*

## **CURRENT STATE OF THE BULGARIAN-ENGLISH PARALLEL CORPUS WITH ALIGNED VERB FORMS**

**TODOR LAZAROV**

INSTITUTE FOR BULGARIAN LANGUAGE, BULGARIAN ACADEMY OF SCIENCES  
*tdlazarov@gmail.com*

This paper deals with the current state of the Bulgarian-English parallel corpus with aligned verb forms. First, I present the objectives for the construction of the corpus and its characteristics in the context of contemporary theory and practice. A description of the linguistic annotation and structural features of the corpus is given. I also present the features of the text types and the primary language data extracted from the corpus.

*Keywords:* corpus, linguistic resource, temporal system, verb forms

### **Българско-английският паралелен корпус със съотнесени глаголни форми в контекста на целите и предназначението му**

Българско-английският паралелен корпус със съотнесени глаголни форми е конструиран през 2017–2018 година, като целта му е да бъде средство за изследване при превод на граматичните паралели между глаголните системи на български и английски, формообразователните способности и значението на граматичните категории, влизаци в състава на темпоралните системи на двата езика. Корпусът се появява като стъпка към формалното описание на граматичните сходства между темпоралните системи на български и английски и извличането на статистически данни за отношенията на преводните реализации на глаголните форми в двата езика.

В контекста на изследванията по темата за граматическите сходства и различия между българската и английската глаголна система предназначението на Българско-английския паралелен корпус със съотнесени глаголни форми е, първо, да представя надеждна, релевантна и адекватна езикова информация и второ, да бъде практически приложим ресурс за създаването на статистически модел за превод на глаголните форми.

Целите на корпуса могат да се дефинират в две основни насоки – теоретични и практически. Теоретичните цели на корпуса са да даде обективни количествени и дистрибутивни езикови данни за глаголните системи на български и английски и да представи изчерпателно (за обема си) качествени и количествени езикови проявления за надежден анализ. Практическите цели на корпуса се свеждат до предоставяне на достатъчно надеждни лингвистични данни за създаването на различни езикови приложения за компютърна обработка на естествения език като статистически модел за аотиране и статистически модел за превод на глаголните форми.

### **Характеристики на Българско-английския паралелен корпус със съотнесени глаголни форми**

От съвременен гледище определението за езиков корпус е по-скоро описателно, отколкото дефинитивно. Определения за езиков корпус се дават чрез неговите характеристики, критерии, цели, както и чрез възможностите, които предоставя, за изследване на даден езиков феномен. Определението, възприето и в този текст, е „колекция от езикови примери, които са избрани и подредени според експлицитни лингвистични критерии, за да се използват като модел на езика" (EAGLES 1996). Проблемни са и строго присъщите характеристики и критерии при определянето на езикови корпуси. Характеристиките на даден езиков корпус могат да бъдат неговата големина, текстова типология, реализацията на езиковите феномени (общи или специализирани), вътрешна структура, анотационна структура и др. В това проучване ще се опитаме да характеризираме Българско-английския паралелен корпус със съотнесени глаголни форми чрез традиционно възприетите характеристики, които до голяма степен представляват синкретично обединение на по-широкия спектър от характеристики за езиков корпус – балансираност, представителност и илюстративност. Трябва да направим уточнението, че и тези характеристики нямат единно тълкуване и принципи за дефиниране, използването им е продиктувано по-скоро от стремеж към концептуализирано прилагане на различен набор от методики при изграждането на езикови корпуси.

Представителността на езиковия корпус се определя със степента, в която той може да обхваща пълния диапазон от проявления на даден езиков феномен. Смята се, че корпусът е представителен, ако констатациите за езика въз основа на неговото съдържание могат да бъдат възприети като обобщения, характеризиращи този език (Leech 1991: 27). Всеки езиков корпус трябва да представлява представителна извадка на езика. Подборът на представителни текстове е основна част от съставянето на всеки корпус за естествените езици. В този смисъл представителността на повечето корпуси до голяма степен се определя от два фактора: обхват на жанрове, включени в корпуса (балансираност), и критериите за подбор на текстове от всеки жанр (илюстративност). Представителността на даден езиков корпус се определя от това как корпусът е балансиран, а балансираността зависи от

текстовата типология – класификацията и характеристиките на включените текстови типове. Балансираният корпус трябва да покрива в някаква пропорция колкото е възможно повече типове текстове, за да бъде качествено и количествено представителен за езика като цяло или за даден езиков феномен. Известни са различни начини за постигане на балансираност, основани както на външни, така и на вътрешни критерии. Необходимо е използваните критерии да бъдат достатъчно разграничими, ефикасни и ограничени на брой, за да се осигури ясното им и еднозначно прилагане при изграждането на балансиран корпус (Коева/Коева 2014). Илюстративността на даден езиков корпус е обвързана с представителността и балансираността му. Определянето на текстовите типове, които да бъдат включени при изграждането на корпуса, налага използваните корпусни единици да демонстрират в максимална степен всички езикови свойства за текстовия тип, към който принадлежат. Постигането на илюстративност е зависимо от избора на представителни текстове, чийто брой за всяка текстова категория трябва да бъде съобразен с честотата или значимостта на тази категория за езиковия феномен, който илюстрират.

Както беше посочено по-горе, Българско-английският паралелен корпус със съотнесени глаголни форми се появява като средство за изследване на граматичните паралели между темпоралните системи на български и английски. От казаното следва, че представителността като характеристика на корпуса е предварително експлицитно зададена цел. Основната цел на корпуса е да бъде средство за надеждно представяне на търсения езиков феномен във всичките му проявления, което оказва влияние върху останалите характеристики на корпуса. Представителността по отношение на преводните реализации на глаголните форми между български и английски поражда конфликт между балансираността и илюстративността на корпуса. По време на работата по създаването му се наложи редица предварително зададени критерии за текстовата типология и за структурата да бъдат модифицирани с цел да се осигури представителност на корпуса по отношение на преводните реализации на глаголните форми.

Важно е да бъде направено уточнението, че Българско-английският паралелен корпус със съотнесени глаголни форми представлява извадки от други езикови ресурси, подбрани по вътрешни критерии. Стъпката, която предхожда създаването на самия корпус, е подборът и оценката на подходящи и надеждни езикови ресурси, които отговарят на целите на корпуса. Оценката на всеки един ресурс е направена според няколко критерия. При първоначалния дизайн на корпуса е възприет един от основните методи за постигане на балансираност при изграждането на езиков ресурс – базирането му на вече възприета структура и извличането на стратифицирани случайни извадки от по-голям корпус (Leech 1991). При първоначалната оценка на подходящите ресурси е избран Българско-английският паралелен корпус със съотнесени изречения (Коева et al. 2012), който е съставна част

от Българско-английския паралелен корпус, който от своя страна е част от Българския национален корпус.

Обемът на Българско-английския паралелен корпус със съотнесени изречения е около 367 000 думи, разпределени неравномерно между български и английски. Корпусът включва различни нива на едноразична и многоразична анотация. По външни критерии са подбрани и ограничен брой текстове от интернет, както и специално конструирани за целта изречения с цел нормализиране на текстовата популация в корпуса. Първоначалният критерий за предварително дефинирана посока на превод на текстовете, включени в корпуса, е изоставен поради липсата на достатъчно представителни оригинални български текстове. Подбраните текстове най-напред са разделени по област и жанр, след което поетапно са включени в състава на Българско-английския паралелен корпус със съотнесени глаголни форми. За всяка една част от подбраните ресурси са извлечени статистически данни за количеството и разнообразието на глаголните форми за дадения текстов тип. След всяко ново добавяне на езиков материал е следено съотношението между общия брой на българските глаголни форми и разнообразието между отделните темпорални форми с цел да се избегне експоненциалното нарастване на отделните форми. В този смисъл таксономията на текстовите типове в корпуса няма единна структура, а е основана на периодична проверка за езиково насищане на корпуса – при внасянето на поредния обработен текст към него се проверява дали съотношението на изследваните езикови единици (глаголните форми) ще се промени значително, или броят им ще остане сравнително постоянен. Както е известно (Viber 1993) езиковите характеристики с висока честота на употреба са сравнително стабилни при разпространението си, докато за по-редките характеристики са необходими големи и предварително подбрани текстове. Практически нулевата фреквентност на някои темпорални форми в български (Куцаров/Kutsarov 2007) предполага, че при фазите на добавяне на езиков материал към корпуса и достигането на езиково насищане определените предполагаеми стойности за честотата на тези темпорални форми няма да се променят значително и те ще запазят ниските си стойности през целия процес на работата. С други думи, при включването на специално подбран по вътрешни критерии текстов тип, който внася нови уникални срещания на темпорални форми към масива, не може да се очаква, че броят на новите срещания ще продължи да се изменя пропорционално при следващите етапи на работа. В този смисъл Българско-английският паралелен корпус със съотнесени глаголни форми представлява не-балансиран езиков корпус поради факта, че външните критерии за текстовите типове са дефинирани на базата на вътрешни езикови критерии.

### **Лингвистична анотация и структура на корпуса**

Лингвистичните данни в корпуса имат два слоя анотация. Първият слой е слой на морфологичните характеристики на думите и се състои от ПоС

таговете на езиковите единици. Възприета е анотационната структура на Българско-английския паралелен корпус със съотнесени изречения поради факта, че текстовете от него представляват по-голямата част от Българско-английския паралелен корпус със съотнесени глаголни форми. Текстовете в Българско-английския паралелен корпус със съотнесени изречения са анотирани с помощта на Българската многокомпонентна система за първична обработка и лингвистична анотация на текстове (Koeva & Genov 2011). За езиковите ресурси, при които няма налична първична анотация, е избран свободно достъпният инструмент за анотиране ТриТагър [TreeTag ger] (Schmid 1995). ТриТагър е инструмент за анотиране на текст с информация за морфологичните характеристики на думите. За да се осигури съвместимостта на анотацията между текстовете от Българско-английския паралелен корпус със съотнесени изречения, които вече притежават анотационен слой, и текстовете, събрани от интернет и тагирани с ТриТагър, след процеса на анотация на текстовете с ТриТагър е приложено конвертиране на таговете с помощта на недетерминиран краен преобразувател, разработен за целта.

Вторичният анотационен слой съдържа информация за глаголните форми в двата работни езика. При работата върху него е спазен принципът на надграждането между равнищата (Koeva et al. 2010: 3681) – различните равнища на анотацията приписват нова информация към езиковите единици, без да променят анотациите от предходните равнища. За втория слой на анотацията е използван свободно достъпният инструмент за работа с езикови ресурси УебАно [WebAnno] (Yimam et al. 2014). В завършения си вид Българско-английският паралелен корпус със съотнесени глаголни форми е записан в преработена версия на файловия формат TONLL-X (Buchholz & Marsi 2006). Анотационните характеристики се кодират в нормализирани текстови файлове (UTF-8, с използване на LF символ за прекъсване на реда, включващ LF символ в края на файла/текста) с три типа редове: редове, съдържащи анотацията на думата/токъна с 5 колони, разделени с табулация; празни редове, маркиращи края на изреченията; и редове с коментари. Изреченията се състоят от един или повече редове с думи, а редовете с думи съдържат следните колони:

- пореден номер на изречението;
- думата/токъна;
- ПоС таг от първия анотационен слой;
- таг на съставната глаголна форма, ако токът е част от такава;
- структурно отношение между двата слоя анотация – представлява I/O структура на частите (chunk I/O structure), съдържа атрибутите В (Begin [начало]), I (In [вътре]/част от) и О (Out [извън]/не е част от).

### **Морфологични характеристики на втория анотационен слой**

При работата върху корпуса са възприети няколко анотационни конвенции. При първичния слой са използвани вече съществуващи инвентари

от ПоС тагове, които са унифицирани в процеса на работа. Лингвистичната анотация на втория анотационен слой в Българско-английския паралелен корпус със съотнесени глаголни форми е разработена специално за целите на корпуса. Инвентарът на втория анотационен слой се състои от 26 стойности, които представят информация за морфологичната категория *време*, разглеждана според класическите наименования на членовете ѝ. При разработването и прилагането на таговете от този слой са възприети следните принципи:

- лингвистичните маркери за български и английски представят традиционно възприетите членове на категорията *време* в двата езика – 9 за български и 17 за английски (16 за темпоралните форми и 1 за инфинитивната форма);
- лингвистичните маркери не съдържат информация за лице, число или род на глаголните форми, тъй като тази информация се формира от стойностите на слоя с ПоС тагове;
- допустимо е съотнесените форми да не бъдат поравно разпределени в двата езика – т.е. възможно е глаголна форма в български да няма съответния еквивалент в английски (преводът да бъде осъществен с други части на речта) и обратно;
- маркерите отразяват темпоралните характеристики на формите за наклонения, страдателен залог и евиденциалните форми, без да отразяват другите морфологични характеристики на тези форми.

Лингвистичните маркери от втория анотационен слой са следните:

- За български:
  - Vpraesens – глаголна форма в сегашно време;
  - Vaor – глаголна форма в минало свършено време;
  - Vimperf – глаголна форма в минало несвършено време;
  - Vfutur – глаголна форма в бъдеще време;
  - Vperfect – глаголна форма в минало неопределено време;
  - Vplusqperf – глаголна форма в минало предварително време;
  - Vfutexact – глаголна форма в бъдеще предварително време;
  - Vfutexpraet – глаголна форма в бъдеще предварително време в миналото;
  - Vfutraet – глаголна форма в бъдеще време в миналото.
- За английски:
  - Vprs – глаголна форма в сегашно просто време;
  - Vps – глаголна форма в минало просто време;
  - Vfs – глаголна форма в бъдеще просто време;
  - Vprp – глаголна форма в сегашно перфектно време;
  - Vpp – глаголна форма в минало перфектно време;
  - Vfp – глаголна форма в бъдеще перфектно време;
  - Vpsc – глаголна форма в сегашно продължително време;
  - Vpc – глаголна форма в минало продължително време;
  - Vfc – глаголна форма в бъдеще продължително време;

- Vprpc – глаголна форма в сегашно перфектно продължително време;
- Vppc – глаголна форма в минало перфектно продължително време;
- Vfpc – глаголна форма в бъдеще перфектно продължително време;
- Vfsp – глаголна форма в бъдеще просто време в миналото;
- Vfpp – глаголна форма в бъдеще перфектно време в миналото;
- Vfcp – глаголна форма в бъдеще продължително време в миналото;
- Vfpcp – глаголна форма в бъдеще перфектно продължително време в миналото;
- Vinf – инфинитив.

#### **Актуално състояние**

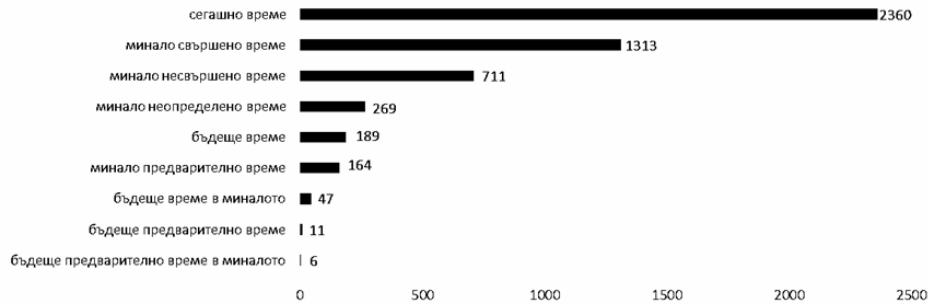
В съвременния си вид Българско-английският паралелен корпус със съотнесени глаголни форми представлява неголям, небалансиран, двуезичен паралелен корпус с общ обем от 5070 аотирани и съотнесени глаголни форми и техните преводни съответствия. Българските текстове съставляват 45 843 думи (48,12% от общия обем), а английските – 49 432 думи (51,88% от общия обем).

Схематично представени, текстовете в Българско-английския паралелен корпус със съотнесени глаголни форми изглеждат така:

- по област:
  - 45% художествени текстове – разкази, романи, приказки и др.;
  - 25% журналистически текстове – новини, изявления, съобщения;
  - 15% научни текстове – изследвания, статии;
  - 10% филмови субтитри;
  - 5% други текстове;
- по посока на превода:
  - 40,1% от текстовете представляват български оригинал с английски превод;
  - 35,4% от текстовете представляват английски оригинал с български превод;
  - 24,5% от текстовете представляват преводи на български и английски от оригинал на трети език;
- по произход:
  - 84% от текстовете представляват текстове от българско-английския паралелен корпус със съотнесени изречения;
  - 15,4% от текстовете са текстове, извлечени от интернет;
  - 0,6% от текстовете са конструирани изречения, извлечени от примери от учебник по граматика за български език.

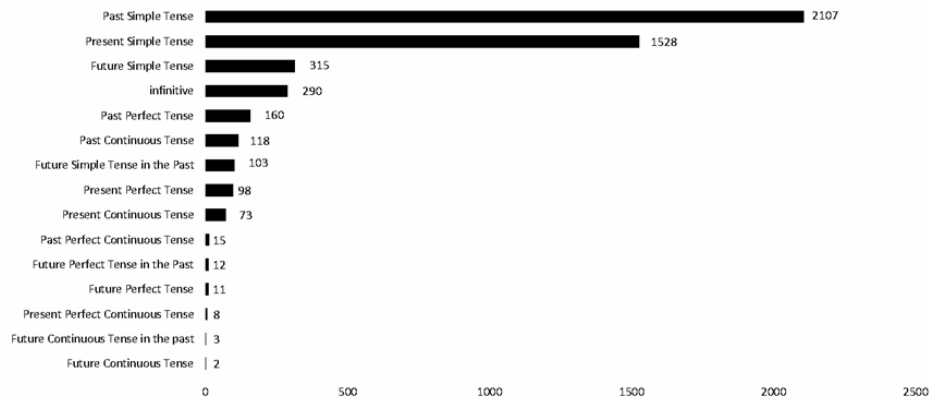
Според разпределението на темпоралните форми в български и английски езиковите данни от Българско-английския паралелен корпус със съотнесени глаголни форми са представени във фигури 1 и 2.

Разпределение на темпоралните форми за български



Фигура 1. Разпределение на темпоралните форми за български език

Разпределение на темпоралните форми за английски



Фигура 2. Разпределение на темпоралните форми за английски език

### Бъдеща работа

Бъдещата работа по корпуса е в две основни направления – реализацията на корпуса като продукт и извличането на статистически данни за граматичните паралели между темпоралните системи в български и английски. За да бъде пълноценен езиков ресурс, е необходимо той да се публикува в интернет и да бъде осигурен свободен достъп до него. Осъществяването на свободен достъп до корпуса ще предостави възможност за бъдещи изследвания върху езиковия материал, които са извън обхвата на първоначално дефинираните за настоящото проучване. Извличането на статистически данни за граматичните паралели между темпоралните системи на български и английски е първоначално поставената цел при създаването на корпуса. Предстои да бъдат направени детайлни изследвания върху отношенията между морфологичните категории в двата езика,



взаимоотношенията между тях, структурните и повърхнинните им реализации при превод.

### **Благодарности**

Тази статия е част от проект № 72-00-40-221/10.05.2017: „Българо-английски граматични паралели с оглед на машинния превод. Обогаляване на статистически модел за превод от български на английски с лингвистична информация“, осъществен с финансовата подкрепа на Програма за подпомагане на млади учени и докторанти на БАН – 2017.

### ЛИТЕРАТУРА

- Коева 2014: *Коева, Св.* Българският национален корпус в контекста на световната теория и практика – В: *Езикови ресурси и технологии за български език*. София, АИ „Проф. Марин Дринов“, с. 29-53.
- Куцаров 2007: *Куцаров, Ив.* Теоретична граматика на българския език. Морфология. Пловдив, УИ „Паисий Хилендарски“.
- Biber 1993: *Biber, D.* Representativeness in Corpus Design. – *Literary and Linguistic Computing*, vol. 8, issue 4, 1 January 1993, pp. 243-257.
- Buchholz & Marsi 2006: *Buchholz, S. and Marsi, E.* CoNLL-X Shared Task on Multilingual Dependency Parsing. – In: *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X '06)*. Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 149-164.
- EAGLES 1996: *Expert Advisory Group for Language Engineering Standarts Preliminary Recommendations on Corpus Typology*. EAG-TCWG-CTYP/P. Pisa, Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale. <<http://www.ilc.cnr.it/EAGLES/corpusstyp/corpusstyp.html>>
- Koeva & Genov 2011: *Koeva, S., A. Genov.* Bulgarian Language Processing Chain. – In: *Proceedings of Integration of Multilingual Resources and Tools in Web Applications*. Workshop in conjunction with GSCL 2011, 26 September 2011, University of Hamburg.
- Koeva et al. 2010: *Koeva Sv., D. Blagoeva, S. Kolkovska.* Bulgarian National Corpus Project. – In: *Proceedings of LREC-2010*, Valletta, ELRA, 3678-3684.
- Koeva et al. 2012: *Koeva, S., B. Rizov, E. Tarpomanova, Tsv. Dimitrova, R. Dekova, Iv. Stoyanova, S. Leseva, Hr. Kukova, and A. Genov.* Application of Clause Alignment for Statistical Machine Translation. – In: *Proceedings of SSST-6, Sixth Workshop on Syntax, Semantics and Structure in Statistical Translation*. Jeju, Republic of Korea, 12 July 2012, The Association for Computational Linguistics, ACL 2012 / SIGMT / SIGLEX Workshop, 2012, pp. 102-110.
- Leech 1991: *Leech, G.* The State of the Art in Corpus Linguistics. – In: *Aijmer, K., B. Altenberg* (eds.). *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. London, Longman, pp. 8-29.
- Schmid 1995: *Schmid H.* Improvements in Part-of-Speech Tagging with an Application to German. – In: *Armstrong S., Church K., Isabelle P., Manzi S., Tzoukermann E., Yarowsky D.* (eds.) *Natural*

Language Processing Using Very Large Corpora. Text, Speech and Language Technology, vol 11. Dordrecht, Springer, pp. 13-22.

Yimam et al. 2014: *Yimam, S., Castilho, E., Gurevych, I.* Automatic Annotation Suggestions and Custom Annotation Layers in WebAnno. – In: *Proceedings of ACL-2014*, demo session.

## REFERENCES

- Biber 1993: *Biber, D.* Representativeness in Corpus Design. – *Literary and Linguistic Computing*, vol. 8, issue 4, 1 January 1993, pp. 243-257
- Buchholz & Marsi 2006: *Buchholz, S. and Marsi, E.* CoNLL-X Shared Task on Multilingual Dependency Parsing. – In: *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X '06)*. Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 149-164.
- EAGLES 1996: *Expert Advisory Group for Language Engineering Standarts Preliminary Recommendations on Corpus Typology*. EAG-TCWG-CTYP/P. Pisa, Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale. <<http://www.ilc.cnr.it/EAGLES/corpusyp/corpusyp.html>>
- Koeva 2014: *Koeva, Sv.* Balgarskiyat natsionalen korpus v konteksta na svetovnata teoriya i praktika [The Bulgarian National Corps in the Context of World Theory and Practice]. – In: *Ezikovi resursi i tehnologii za balgarski ezik*. Sofia, AP „Prof. Marin Drinov”, pp. 29-53.
- Koeva & Genov 2011: *Koeva, S. & A. Genov*, Bulgarian Language Processing Chain. – In: *Proceedings of Integration of Multilingual Resources and Tools in Web Applications*. Workshop in conjunction with GSCL 2011, 26 September 2011, University of Hamburg.
- Koeva et al. 2010: *Koeva Sv., D. Blagoeva, S. Kolkovska*. Bulgarian National Corpus Project. – In: *Proceedings of LREC-2010*, Valletta, ELRA, 3678-3684.
- Koeva et al. 2012: *Koeva, S., B. Rizov, E. Tarpomanova, Tsv. Dimitrova, R. Dekova, Iv. Stoyanova, S. Leseva, Hr. Kukova, and A. Genov*. Application of Clause Alignment for Statistical Machine Translation. – In: *Proceedings of SSST-6, Sixth Workshop on Syntax, Semantics and Structure in Statistical Translation*. Jeju, Republic of Korea, 12 July 2012, The Association for Computational Linguistics, ACL 2012 / SIGMT / SIGLEX Workshop, 2012, pp. 102-110.
- Kutsarov 2007: *Kutsarov, Iv.* Teoretichna gramatika na balgarskiya ezik. Morfologiya [Theoretical Grammar of the Bulgarian Language. Morphology]. Plovdiv, UI „Paisiy Hilendarski”.
- Leech 1991: *Leech, G.* The State of the Art in Corpus Linguistics. – In: *Aijmer, K., B. Altenberg* (eds.). *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. London, Longman, pp. 8-29.
- Schmid 1995: *Schmid H.* Improvements in Part-of-Speech Tagging with an Application to German. – In: *Armstrong S., Church K., Isabelle P., Manzi S., Tzoukermann E., Yarowsky D.* (eds.). *Natural Language Processing Using Very Large Corpora. Text, Speech and Language Technology, vol 11*. Dordrecht, Springer, pp. 13-22.

Yimam et al. 2014: *Yimam, S., Castilho, E., & Gurevych, I.* Automatic Annotation Suggestions and Custom Annotation Layers in WebAnno. – In: *Proceedings of ACL-2014*, demo session.

### РЕЗЮМЕ

Текстът представя съвременното състояние на Българско-английския паралелен корпус със съотнесени глаголни форми. Представени са целите, за които е създаден корпусът, характеристиките на корпуса в контекста на съвременната теория и практика. Дава се описание на лингвистичната анотация и структурните особености на корпуса. Представени са особеностите на текстовите типове и първични езикови данни, извлечени от корпуса.

*Ключови думи:* корпус, езиков ресурс, темпорална система, глаголни форми

✉ *Тодор Лазаров*

Секция по компютърна лингвистика

Институт за български език „Проф. Л. Андрейчин“ при БАН

бул. „Шипченски проход“ 52, бл. 17, 1113 София, България

✉ *Todor Lazarov*

Department of Computational Linguistics

Institute for Bulgarian Language, Bulgarian Academy of Sciences

52 Shipchenski prohod, Bl. 17, 1113 Sofia, Bulgaria

Публикувано: 31 март 2019